

Title	観光ガイドシステムに必要な知識のWeb文書からの自動獲得
Author(s)	柿澤, 康範
Citation	
Issue Date	2009-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/8123
Rights	
Description	Supervisor:東条 敏, 情報科学研究科, 修士

Automatic acquisition of knowledge for sightseeing guide systems

Yasunori Kakizawa (710017)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 5, 2009

Keywords: attribute-value, trouble, web document, large corpora.

In this thesis, we describe automatic classification methods of attribute-value and trouble information of a given topic. The classification methods were designed for fulfilling user's needs in sightseeing and the resulting knowledge is to be used in spoken dialog systems that guide tourists in Kyoto. More specifically, the goal of this paper is to associate a user's action ("go", "see", etc.) in sightseeing to particular types of information presented in the form of attribute-values and troubles that are automatically acquired from a huge document collection in the Web. We attempted 1) to classify attributes according to a user's action such that the action demands users to know the information about the values of the attributes and 2) to classify nouns expressing troubles according to their seriousness, which are represented by a ranked list of typical verbs expressing troubles. According to the classification of troubles, dialog systems can provide the information concerning relatively small number of the troubles that interfere with particular actions to be taken taken by sightseers among many other troubles. Experimental results showed 1) that the accuracy of the resulting associations between attributes and actions was around 42%, and 2) that the classification of trouble-nouns was done with about 84% accuracy. We also tried to judge the degree of seriousness of troubles by automatically deciding which one of given two trouble nouns are more serious. The accuracy of this judgement was 68% (2classes-classification was

about 97%). As future work, we are going to use the obtained knowledge on attribute-values and troubles in real-world spoken dialog systems in the next year.

In this thesis, the words referring to those aspects of an object in which people are interested are called attributes. For instance, if the object is a “寺 (temple)”, “拝観料 (admission fee)” and “交通手段 (access)” should be regarded as attributes. Attributes typically have values. The admission fee of a temple should have a value such as “300 yen”. Yoshinaga et al.?? proposed a method to automatically acquire such pairs of attributes and values of given objects from Web documents. We attempt to classify attributes relating to a user’s action (e.g., “go”, “see”) such that the action demands users to know the value of the attributes. Using this knowledge, the sightseeing guide system can cater to user’s needs. For instance, if a user wants to go to Kiyomizu temple, the values of the attributes “交通手段 (access) ” and “住所 (address)” are crucial. On the other hand, if a user wants to see Kiyomizu temple, the values of the attributes “見所 (highlights)” should be taken into account. We developed a classification method for attributes and evaluated it. Experimental results showed that the accuracy of the proposed method was 42%, which is 16% better than a baseline method.

We also attempted to classify nouns referring to troubles. We called such nouns trouble nouns. Examples of trouble nouns are “渋滞 (traffic jam)” and “人混み (crowds)”. The classification method categorizes the trouble nouns into classes each of which corresponds to verbs also expressing troubles such as “遅れる (delay)” and “疲れる (being exhausted)”. These verbs are useful for recognizing the seriousness of troubles and for identifying the troubles that are important for sightseers, and we are going to use the classification results to detect troubles to be presented to sightseers by the dialog system for sightseeing in Kyoto. Experimental results showed that our classification method achieved around 84% accuracy.

In addition, we developed a method that assesses the seriousness of troubles by the above classification results based on Scheffe paired comparison (5-class classification). We manually prepared training data and applied machine learning techniques (Support Vector Machine, Maximum Entropy Method) for paired comparison. Experimental results showed that the ac-

curacy of SVM was 65% and the accuracy of the MEM was 68% (with 2-class classification by specific pairs around 97%). As future work, we will improve the accuracy of the proposed methods and acquire knowledge about a user's potential action plans. In the next year we plan to use the acquired knowledge on attribute-values and troubles in a real-world spoken dialog system for electronic sightseeing guides in Kyoto.

References

- [1] Naoki Yoshinaga and Kentaro Torisawa, "Open-Domain Attribute-Value Acquisition from Semi-Structured Texts" In Proceedings of the Workshop on Ontolex 2007 – The Lexicon/Ontology Interface held at the sixth International Semantic Web Conference, pp. 55-66. Nov., 2007.
- [2] S. De Saeger, K. Torisawa, and J. Kazama. Looking for trouble. In Proc. of The 22nd International Conference on Computational Linguistics (Coling2008), 2008.